Deep Kalman Filters

Time Series Forecasting of Sales

C. Saharia	150050011
H. Goka	150050069
V. Sreeramdass	150050084
V. Singh	150050046

Motivation

Variational models provide a robust mechanism to capture patterns in data without overfitting.

They are however challenging to train and use.

The motivation lay in learning about these models in a temporal setting.

Specifically the Rossman Sales Dataset, for which Deep Kalman Filters are the intuitive choice.

Literature Survey

Auto-Encoding Variational Bayes - D. Kingma and M. Welling

Provides foundation for the model to be extended to capture temporal patterns

Deep Kalman Filters - Krishnan et al.

The main reference for the model, applied in a different setting

Approaches tried so far...

Baseline

LSTM Networks - These models have enjoyed remarkable success in the past few year to capture the dependency across time steps.

- Basic LSTM modified to feed the predicted values to the next time step
- Training with Teacher Forcing
- Training with Scheduled Sampling

Deep Kalman Filters

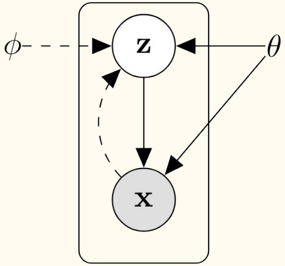
Representation of the current state of the system in latent space

Progression of latent space by *transition model*

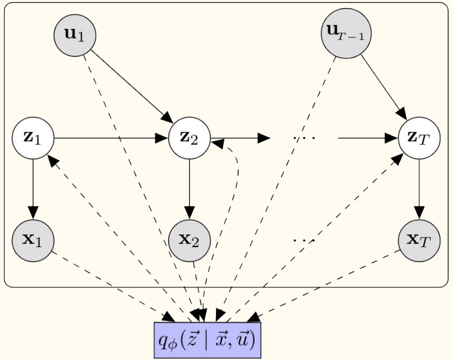
Generation from latent space by *generation model*

Use of a *recognition network* to train

the generation and transition models



(a) Variational Autoencoder



(b) Deep Kalman Filter

Performance

Model	Training Error	Validation Error	RMSPE
Basic LSTM	0.12	0.19	0.24
LSTM Teacher Forcing	0.08	0.1	0.29
LSTM Scheduled Sampling	0.07	0.11	0.27
Deep Kalman Filters	-	-	-

Failures ⇔ Lessons

- Too many parameters to train
  - Use of clever embeddings
- Unstable loss function and exploding gradients
  - Handling rounding errors
  - Gradient clipping
- Sparse dataset with very long term dependencies
  - Large latent space
- Complex Variational Architecture

Demonstration